Minimizing Noise in Location Privacy Protection Through Equipment Error Consideration

Original Scientific Paper

Riho Isawa
The University of Electro-Communications,
Graduate School of Informatics and Engineering
Departments, Department of Informatics
1-5-1 Chofugaoka, Chofu, Japan
isawa.riho@ohsuga.lab.uec.ac.jp

Yasuyuki Tahara
The University of Electro-Communications,
Graduate School of Informatics and Engineering
Departments, Department of Informatics
1-5-1 Chofugaoka, Chofu, Japan
tahara@uec.ac.jp

Yuichi Sei
The University of Electro-Communications,
Graduate School of Informatics and Engineering
Departments, Department of Informatics
1-5-1 Chofugaoka, Chofu, Japan
seiuny@uec.ac.jp

Akihiko Ohsuga
The University of Electro-Communications,
Graduate School of Informatics and Engineering
Departments, Department of Informatics
1-5-1 Chofugaoka, Chofu, Japan
ohsuga@uec.ac.jp

Abstract – *In recent years, systems that collect location information and publish statistics, such as those that publish congestion information, have been extensively employed. Because it is possible to infer an individual's identity even if the information is not directly disclosed, it is essential to disclose data with privacy protection. Therefore, privacy protection methods based on differential privacy are attracting attention. Geo-indistinguishability is the most famous extension theorem of differential privacy for location information. Geo-indistinguishability can be achieved by adding noise to a target value that must be protected. However, noise addition reduces the usefulness of the data. Thus, it is desirable to add minimal noise to your privacy budget. Therefore, we focus on the fact that the values obtained using measurement devices contain errors. We introduced a novel concept of differential privacy tailored for location information, termed true-value-based geo-indistinguishability ($T\text{-}Geo\text{-}I$), which accounts for equipment noise. We also proposed a location information privacy protection method that considers $T\text{-}Geo\text{-}I$ and reduces the amount of added noise. The object of privacy protection should be the "true value" not the "measured value" that includes measurement errors.*

According to the experimental results, in the case wherein the measurement error is the normal distribution, our method reduced the noise average and mean square error (MSE) by up to 41% and 63%, respectively, compared with conventional methods while maintaining a prespecified level of privacy in 10^8 samples of numerical data. In the case wherein the measurement error is the lognormal distribution, the proposed method based on $T\text{-}Geo\text{-}I$ succeeded in reducing the noise average and MSE by up to 60% and 67%, respectively, compared with methods based on $Geo\text{-}I$, while maintaining a prespecified level of privacy. These findings indicate that the proposed method can improve the usefulness of data while maintaining a prespecified degree of privacy protection.

Keywords: *differential privacy, geo-indistinguishability, GPS, Laplace mechanism, local differential privacy, location-based service*

Received: December 12, 2023; Received in revised form: February 10, 2024; Accepted: February 10, 2024

1. INTRODUCTION

In recent years, systems that collect location information and publish statistics, such as those that publish congestion information, have been extensively employed. The Internet of Things (IoT) technology has revolutionized innovation in people's lives by collecting and storing information received from physical objects or sensors [1–2]. Although these systems are convenient, they carry the risk of leaking personal information such as location information [3]. Even if personal information is not directly disclosed, it may be inferred from statistical data. Storing and using information on the cloud is also becoming more prevalent [4–5]. Location privacy preservation is essential, and there are many research challenges [6–7].

When disclosing statistical data to the public, it is essential to take privacy into account and perform processing to ensure that individuals cannot be identified from the data before releasing the data. Recently, privacy protection methods based on differential privacy have attracted attention. Representative examples of privacy protection for location information based on differential privacy include NTT Docomo's mobile spatial statistics and Google Maps processing of congested areas. Differential privacy is used in statistics in the real world and is widely recognized as a security indicator that can suppress the disclosure of data privacy, regardless of the attacker's background knowledge or attack method algorithm.

Geo-indistinguishability (*Geo-I*) is attracting attention as a standard that applies differential privacy to protect location information data [8]. It shows the guaranteed criteria when noise is added to the position information using the perturbation method on the Euclidean plane. One perturbation method that satisfies the *Geo-I* criteria and protects the true value by adding random noise to a person's location information is the planar Laplace mechanism. This method protects privacy by adding noise that satisfies the criterion of differential privacy to the true data using the Laplace distribution. In general, the stronger the degree of privacy protection, the higher the amount of noise added, which reduces the usefulness of the data. There is a trade-off between the usefulness of data and the degree of privacy protection.

Because the degree of privacy protection is specified numerically, there is a need for a noise addition method that satisfies this degree of protection in terms of differential privacy. To enhance the usefulness of the data, the amount of noise added to the true value should be reduced. The more noise added, the less useful the data becomes. Therefore, we focus on the fact that the measured values already contain errors and attempt to suppress the total amount of added noise. Because conventional methods do not consider errors during measurement, they may contain extra noise for the privacy protection parameter budget. In general, technologies for obtaining location information include GPS, Wi-Fi, beacons, and communication base stations. Because it is measured using IoT equipment, it already contains errors. To maintain a prespecified degree of privacy protection and enhance the usefulness of the data, we propose a method for reducing the total amount of added noise by considering errors already included in the measured values.

The principal contributions of this study are threefold. First, we introduce a novel concept of differential privacy tailored for location information, termed true-value-based *Geo-I* (*T-Geo-I*), which accounts for equipment noise. Second, we devise an anonymization algorithm that adheres to the *T-Geo-I* standard. Third, we demonstrate that the proposed *T-Geo-I* framework not only upholds the predefined privacy threshold but also reduces noise addition compared with existing methodologies.

The remainder of the paper is organized as follows: Section 2 reviews existing research related to differential privacy. Section 3 defines a new privacy metric and proposes a privacy protection algorithm that ensures compliance with this metric. Sections 4 and 5 detail the experimental method and the results, respectively. Section 6 discusses the experimental results of our proposed method. Finally, Section 7 concludes the study.

2. RELATED WORK

2.1. OVERVIEW OF LOCATION PRIVACY RESEARCH

A significant amount of research has been conducted on location information privacy [9–10]. One famous research field is differential privacy. *Geo-I* is famous for the differential privacy of location information [8].

According to recent research, *Geo-I* in indoor environments has been proposed [11]. The proposed framework introduces two distance calculation and received signal strength (RSS) generation methods based solely on RSS values as novel methods, which have been proven to perform.

Geo-I for task allocation in spatial crowdsourcing has been investigated [12]. An optimized global grouping with the adaptive local adjustment method OGAL with a convergence guarantee was proposed and proven that it works.

These methods do not consider measurement errors; therefore, our method can be applied to make them more efficient to enhance the usefulness of data.

Research on federated learning has been actively conducted recently [13]. Our method can also be incorporated into this. Details are explained in Section 2.10.

2.2. ϵ-DIFFERENTIAL PRIVACY

Differential privacy is extensively used as a strong mathematical definition to protect datasets without relying on attackers' prior information [14–16]. Rather than relying on encryption, differential privacy offers protection by adding noise to the data, and the results are calculated from the data. Because encryption is not involved, the computational cost of differential privacy is low, and it tends to be easy to introduce into many systems.

When mechanism K is a privacy protection function, $S \subseteq Range(K)$, $\epsilon \in R+$, and databases D and D' are adjacent, ϵ-differential privacy is satisfied when the following equation is satisfied. ϵ is a privacy level parameter and a positive number. When the privacy level parameter ϵ is large, the privacy level is low; when ϵ is small, i.e., close to 0, the degree of privacy protection is high. Adjacent means that the records are different in one place. For example, D represents a database with one record removed from D', otherwise D represents a database with one record of D' replaced by another record. This means that D and D' are adjacent.

$$Pr[K(D) \in S] \leq exp(\epsilon) \times Pr[K(D') \in S]. \quad (1)$$

This equation indicates that privacy is protected because different parts of the records cannot be identified if the results from adjacent databases are indistinguishable. For example, if an attacker knows all information except for a certain record A, it is possible to infer the data about A by back-calculating from the database result. Consequently, we can protect privacy by applying for protection according to this guarantee.

2.3. (E,Δ)-DIFFERENTIAL PRIVACY

Differential privacy is mathematically rigorous. It has been mathematically proven that a noise generation method based on the Laplace mechanism using the Laplace distribution has a probability density function ratio of less than the privacy level parameter ϵ in all ranges [17].

For example, for a noise generation method using a normal distribution noise, the ratio of probabilities becomes infinite at the tails of the distribution. Therefore, differential privacy is not guaranteed over the entire region.

However, it is too strict a definition to consider extreme points that seldom occur in reality. (ϵ,δ)-differential privacy allows cases wherein differential privacy is not satisfied if the probability is below a certain level [18].

When mechanism K is a privacy protection function, ϵ is a privacy level parameter, $S \subseteq$ Range(K), $\epsilon \in R+$, and databases D and D' are adjacent, if differential privacy based on the privacy level parameter is not satisfied with a probability less than or equal to δ, Equation 2 is satisfied.

$$Pr[K(D) \in S] \leq exp(\epsilon) \times Pr[K(D') \in S] + \delta. \quad (2)$$

2.4. LOCAL DIFFERENTIAL PRIVACY

The definition of ϵ-differential privacy refers to the protection of the database. Although this is guaranteed for databases that store data, it is not assumed that each data is sent to the server one by one each time. Therefore, the concept of local differential privacy has been proposed [19-20].

When x and x' represent databases of size 1 and protection is performed by mechanism A, for any output y, $\epsilon \in R+$, if Equation 3 is satisfied; for the privacy level parameter ϵ, it satisfies ϵ-local differential privacy.

$$Pr[A(x) = y] \leq exp(\epsilon) \times Pr[A(x') = y]. \quad (3)$$

This standard also allows you to protect your device before sending data to an untrusted server. Therefore, it is possible to collect and use data while protecting the data regardless of the trustworthiness of the server.

2.5. PLANAR LAPLACE MECHANISM

The planar Laplace mechanism is a typical privacy protection method based on differential privacy [17]. It uses the Laplace distribution to generate noise and adds it to the true value to protect privacy. When protecting individual data before sending it to the server, noise is added to each piece of data each time according to local differential privacy before sending it to the server. Because this method differs from encryption, it can protect user privacy with low computational costs. Therefore, it can be easily introduced into many systems. It can be executed on each user's IoT device or smartphone without a significant burden.

However, this method reduces the usefulness of the data. There is a trade-off between the usefulness of data and the degree of privacy protection. Many studies have been conducted to address this disadvantage, and our research is one of them to improve the usefulness of data.

2.6. ϵd_x-PRIVACY

Chatzikokolakis et al. [21] extended differential privacy, which is defined only in databases. $P(Z)$ denotes the probability distribution on Z. K:$X \rightarrow P(Z)$ denotes a mechanism in some domain X that provides a probability distribution in some domain Z. $Dx(x, x')$ is the hamming distance between x and x' on X. ϵ denotes a privacy level parameter, $\epsilon \in R+$, x, $x' \in X$, and $Z \subseteq Z$. If the mechanism K is expressed by Equation 4, ϵdx-privacy is guaranteed.

$$\frac{K(x)(Z)}{K(x')(Z)} \leq \epsilon d_X(x, x'). \quad (4)$$

This definition indicates that the more similar two databases are, the more similar the generated distributions should be.

2.7. GEO-INDISTINGUISHABILITY

Geo-I is a privacy guarantee standard for location information data. It has received particular attention among perturbation methods [22]. *Geo-I* applies ϵd_x-privacy to location information data. It also uses the concept of local differential privacy.

In Equation 5, X represents a set of points of interest, $x,x' \in X$, $d(x,x')$ denotes the distance between x and x' on the Euclidean plane, ϵ denotes a privacy level parameter, $\epsilon \in R+$, Z contains spatial points, and $Z \subseteq Z$. If the mechanism K is expressed by Equation 5, ϵ-*Geo-I* is guaranteed.

$$\frac{K(x)(Z)}{K(x')(Z)} \leq e^{\epsilon d(x,x')}. \quad (5)$$

2.8. PLANAR LAPLACE MECHANISM FOR *GEO-I*

The planar Laplace mechanism is used as a data protection method to satisfy *Geo-I* [22]. This is a method for position information wherein noise is generated from the privacy level parameter ϵ using the Laplace distribution and added to the true position.

For the noise radius value r, we substitute the noise calculated using Equation 6.

$$r_\epsilon(p) = -\frac{1}{\epsilon}\left(W_{-1}\left(\frac{p-1}{\epsilon}\right) + 1\right).$$

For the direction of noise value θ, we randomly calculate a value from the probability of a uniform distribution with [0, 2π). For p, we randomly calculate a value from the probability of a uniform distribution on [0,1), assign it to the true value x, and use $<r\cos\theta, r\sin\theta>$ as noise. We select the closest possible coordinate system to the coordinates with added noise and use that as the value after applying the mechanism. Function W_{-1} denotes Lambert's W function (the −1 branch). This operation guarantees ϵ-*Geo-I*. ϵ denotes a privacy level parameter, and $\epsilon \in R+$.

The probability of obscuring the true position x to x' is calculated using Equation 7. This planar Laplace mechanism rounds the decimal point of the position data. Equation 7 also considers the effect of rounding. d(x, x') denotes the distance between x and x' on the Euclidean plane. ϵ denotes a privacy level parameter.

$$D_\epsilon(x)(x') = \left(\frac{\epsilon^2}{2\pi}\right)e^{-\epsilon d(x,x')}. \quad (7)$$

2.9. TRUE-VALUE-BASED DIFFERENTIAL PRIVACY

Sei et al. [23] proposed the concept of true-value-based differential privacy (TDP). This is a privacy guarantee standard that considers the fact that the values measured using IoT devices contain errors.

The conventional method satisfies the specified degree of privacy protection for the measured value, i.e., "true value + measurement error." However, to meet these criteria, the privacy of the "true value" should be protected with a specified degree of privacy protection. Because noise in the form of measurement errors is already present, the amount of additional noise required to protect privacy is small for the necessary privacy parameter budget compared with the conventional method. Focusing on measurement errors, we attempt to reduce the total amount of noise added according to differential privacy.

For a database of size 1 for x and x', mechanism M is a function that adds error during measurement, and protection is provided by mechanism A. For any output y and $\epsilon \in R+$, when Equation 8 is satisfied, ϵ-differential privacy is satisfied. In addition, TDP assumes that the measurement error is based on a normal distribution.

$$Pr[A(M(x)) = y] \leq exp(\epsilon) \times Pr[A(M(x')) = y]. \quad (8)$$

Considering this concept, even if noise below an appropriate threshold is not added to the measured value, the prespecified degree of privacy protection can be maintained, and the total amount of added noise can be reduced.

TDP concentrates on one-dimensional data [23]. TDP aims to find the optimal maximum w that fulfills Equation 9.

$$e^\epsilon \geq \frac{\mathcal{V}(x + \Delta/2; \sigma^2, \Delta/\epsilon, w)}{\mathcal{V}(x - \Delta/2; \sigma^2, \Delta/\epsilon, w)} \quad (9)$$

where

$$\begin{aligned}\mathrm{V}(\mathrm{x};\ \sigma^2, b, w) &= \textstyle\int_{-\infty}^{\infty} \mathcal{N}(t;\sigma^2)\widehat{\mathcal{L}}(x-t;b,w)dt \\ &+ \mathcal{N}(x;\sigma^2)\textstyle\int_{-w}^{w}\mathcal{L}(t;b)dt \\ &= \frac{e^{-\frac{w+x}{b}-\frac{x^2}{2\sigma^2}}}{4b\sigma} \times \left\{\sigma e^{\frac{1}{2}\left(\frac{2bw+\sigma^2}{b^2}+\frac{x^2}{\sigma^2}\right)}\left[\mathrm{erfc}\left(\frac{b(w-x)+\sigma^2}{\sqrt{2}b\sigma}\right)\right.\right. \\ &\left.\left.+ e^{\frac{2x}{b}}\mathrm{erfc}\left(\frac{b(w+x)+\sigma^2}{\sqrt{2}b\sigma}\right)\right] + 2\sqrt{\frac{2}{\pi}}b\left(e^{\frac{w}{b}}-1\right)e^{\frac{x}{b}}\right\}\end{aligned}$$

and

$$\widehat{\mathcal{L}}(x;b,w) = \begin{cases} \int_{-w}^{w}\mathcal{L}(t;b)dt & x = 0 \\ \frac{e^{-x/b}}{2b} & x \geq w \\ \frac{e^{x/b}}{2b} & x \leq -w \\ 0 & otherwise. \end{cases}$$

Here, ϵ denotes a privacy level parameter, Δ means the range of possible values for numerical attitude, σ means the standard deviation of normal distribution, and b means the scale parameter of Laplace distribution (equal to Δ/ϵ).

The larger the threshold w, the more pronounced the reduction effects. TDP assumes that the measurement error adheres to a one-dimensional normal distribution $N(t;\sigma^2)$. If the measurement error diverges from a one-dimensional normal distribution, a fundamentally different mathematical discussion is required. Even with a one-dimensional normal distribution, as intricate as described by Equation 9, extending Equation 9 to two dimensions is not straightforward.

2.10. COMPOSITION THEOREM FOR HETEROGENEOUS MECHANISMS

Kairouz et al. [24] focused on privacy guarantees under k-fold composition. According to theorem 3.3 in [24], any k-fold adaptive composition of (ε, δ)-differentially private mechanisms satisfies the privacy guarantee. This means that the total privacy budget is obtained during composition.

2.11. FEDERATED LEARNING

Federated learning is a method that protects privacy by training machine learning models on each device [13, 25]. Each local device uses its data to train the model from the central server. Subsequently, only the extracted parameters are aggregated in the central server to improve the accuracy of the common model in the central server.

Federated learning of location information is also being researched [26–27]. For example, population modeling and population density can be estimated without the user having to send the true original data using the proposed method [27]. With federated learning, each device uses data to perform calculations and incorporates them into a machine learning model before sending the data to the server. It is highly compatible with local differential privacy. Our method is based on local differential privacy. There is a high possibility that our proposed method will be incorporated into federated learning to enhance the usefulness of data while maintaining a prespecified privacy protection level.

3. PROPOSED METHOD

3.1. TRUE-VALUE-BASED GEO-I (*T-GEO-I*)

We propose true-value-based geo-indistinguishability (*T-Geo-I*), a privacy protection standard for location information that considers measurement errors. This is a combination of *Geo-I*, which is a privacy protection standard related to location information, and TDP, which is a privacy protection standard that considers measurement errors.

TDP is focused on one-dimensional data. This leads to the meaningful proposition of amalgamating TDP with the privacy protection property of geo-indistinguishability for two-dimensional location information. The challenge in the theoretical analysis of the cumulative effect of measurement errors and differential privacy noise on two-dimensional location data is significant, rendering the direct application of the methodologies proposed in [23] unfeasible. In addition, the research on TDP, as discussed in [23], is confined to scenarios assuming a normal distribution of measurement errors. The uniqueness of the algorithm proposed in Section 3.2 of our study stems from its consideration of cases in which the measurement error does not conform to a normal distribution. This innovative approach significantly extends the applicability and relevance of TDP, particularly in contexts in which data distributions are non-normal. Obtained through simulation, our proposed algorithm is adaptable to any probability distribution. Typically, technologies for acquiring location data encompass GPS, Wi-Fi, beacons, and cellular base stations. Given the variety of devices and the indeterminate nature of measurement error distributions, the versatility of the proposed method in accommodating various error distributions is of substantial significance.

Let mechanism M be a function that adds error during measurement, X is a set of points of interest, $x, x'\in X$, $d(x, x')$ denotes the distance between x and x' on the Euclidean plane, ϵ denotes a privacy level parameter, $\epsilon\in R+$, Z contains spatial points, and $Z\subseteq Z$. ϵ-*T-Geo-I* is guaranteed when mechanism K satisfies Equation 10.

$$\frac{K\big(M(x)\big)(Z)}{K\big(M(x')\big)(Z)} \leq e^{\epsilon d(x,x')}. \tag{10}$$

3.2. PRIVACY PROTECTION METHOD BASED ON *T-GEO-I*

Privacy protection method based on *T-Geo-I* is based on the planar Laplace mechanism. As mentioned in Section 2.4., because the planar Laplace mechanism has a low computational cost to protect privacy and is easy to use in various systems, our method incorporates this mechanism.

We propose a method wherein no noise is added to the data when the noise generated using the planar Laplace mechanism of *Geo-I* is below the threshold w; the noise is added to the data when the noise is the threshold w or above. Noise generation follows Section 2.7. The value generated using Equation 6 is the radius of the noise added to the measurement noise value, and the threshold w determines whether noise is added.

The problem with the proposed method is that it is difficult to solve the threshold value w analytically. In previous research [23], w was determined by calculation using mathematical formulas. We solve this problem by finding the threshold value w through simulation.

The pseudocode for the privacy protection method is shown in Algorithm 1. In the proposed method for analytically adding privacy noise, the noise radius r is calculated using Equation 11. For θ, we randomly calculate a value from the probability of a uniform distribution with [0,2π). For p, we randomly calculate a value from the probability of a uniform distribution on [0,1). ϵ can be any positive value determined as a privacy level parameter.

Because of the proposed method, it is necessary to find an appropriate threshold value w for the noise radius r. The optimal threshold w value is the minimum value within the range that satisfies Equation 10.

Algorithm 2 illustrates the algorithm for determining the optimal threshold w. To confirm that Equation 10 is satisfied, a total noise probability density function is derived by combining the measurement error and privacy noise. Because the probability density function cannot be derived through calculation, it is derived by randomly generating ns samples as an experiment. A probability density function shifted by *Δ* is also derived. Differential privacy is satisfied when the ratio of the two probability density functions satisfies Equation 10. Because the accuracy of the probability density function is low in areas with few samples, only the areas with (1-δ) samples are checked. If differential privacy is satisfied, even with a sufficiently large threshold w, let w be infinite.

MeasurementNoise(), in the 8th line in Algorithm 2., returns the value obtained from the distribution of measurement errors. The distribution of measurement error is not limited to a normal distribution. The noise distribution may be any distribution and can be changed depending on the measuring equipment.

PrivacyNoise() in the 10th line in Algorithm 2. is the algorithm shown in Algorithm 1.

$$r_{\epsilon}(p) = -\frac{1}{\epsilon}\left(W_{-1}\left(\frac{p-1}{\epsilon}\right) + 1\right), \tag{11}$$

$$|PrivacyNoise(\epsilon, w)| = \begin{cases} r_{\epsilon}(p) & (w < r_{\epsilon}(p)) \\ 0 & (r_{\epsilon}(p) < w). \end{cases} \tag{12}$$

Algorithm 1. Privacy protection mechanism for location information considering measurement errors.

```
Input: $\epsilon$ (Privacy level parameter), $v_x$, $v_y$ (Measured location values), $w$ (Threshold value)
Output: TDP value
1: Generate a random value $p$ from a uniform distribution [0,1)
2: $r \leftarrow r\epsilon(p)$
3: Generate a random value θ from a uniform distribution [0, 2π)
4: if $r < w$ then
5:     return ($v_x$, $v_y$).
6: else
7:     return ($v_X + rcos\theta$, $v_y + rsin\theta$).
8: end if
```

Algorithm 2. Algorithm for determining threshold w.

```
Input: $\epsilon$ (Privacy level parameter), $c$ (Width of a histogram), $\Delta$ (Distance of $x$ and $x'$), $\delta$ (Scope of verifying differential privacy), $\alpha$ (Multiple of $w$ to verify), $ns$ (Number of samples)
Output: Threshold $w$ used in the proposed method
1: for $w = \alpha, 2\alpha, ...$ do
2:     isDF ← true
3: {Prepare two array variables as Histogram}
4:     $B \leftarrow b1, b2, \ldots$
5:     $B' \leftarrow b1', b2', \ldots$
6:     for $i = 1, \ldots, ns$ do
7: {Add measurement error}
8:     $v \leftarrow$ MeasurementNoise()
9: {Add Laplace noise considering threshold $w$}
10:    $v \leftarrow v + PrivacyNoise(\epsilon, w)$
11: {Calculate the corresponding bin of the histogram of value $v$.}
12:    $index \leftarrow \lceil |v|/c \rceil$
13:    $b_{\text{index}} \leftarrow b_{\text{index}} + 1$
14: {Calculate the corresponding bin of the histogram of value $|v| + \Delta$.}
15:    $index' \leftarrow \lceil (|v| + \Delta)/c \rceil$
16:    $b_{\text{index}} \leftarrow b_{\text{index}} + 1$
17:    end for
18: {Determine the scope to verify differential privacy}
19:    $sum \leftarrow 0$
20:    $threshold \leftarrow 0$
21:    for i = 1,... B'.length do
22:        $sum \leftarrow sum + b_i'$
23:        if $sum/ns > 1 - \delta$ then
24:            $threshold \leftarrow i$
25:            break
26:        end if
27:    end for
{Verify whether differential privacy is satisfied}
28:    for $i = 1, \ldots,$threshold do
29:        if $b_i/b_i' > \exp(\epsilon\Delta)$ or $b_i'/b_i > \exp(\epsilon\Delta)$ then
30:            $isDF \leftarrow false$
31:            break
32:        end if
33:    end for
{Return value if differential privacy is not satisfied}
34:    if not isDF then
35:        return $w - \alpha$
36:    end if
37: end for
```

4. EXPERIMENT METHOD

4.1. SIMULATION METHOD

We simulated the proposed method. We compared the proposed method *T-Geo-I* with the planar Laplace mechanism for methods based on *Geo-I* [22] and TDP [23].

The simulation was performed in two scenarios. One involved performing experiments by setting a person's position to (0, 0) and adding noise as a numerical simulation. The other involved dividing people into grids and conducting a simulation experiment to count the number of people on each grid.

In the grid experiment, we used data generated using the Siafu simulation tool [28]. The Siafu tool is open-source software for obtaining data on human behavior using a typical human behavior model on a map. The setup includes 10,000 users interacting in a space measuring 8.4 km x 8.4 km, which includes businesses, restaurants, and parks. We used the data for this simulation based on previous research by Sei et al. [29].

In this experiment, the measurement error assumes 2 types, a normal distribution and a lognormal distribution. MeasurementNoise(), in the 8th line in Algorithm 2., returns noise based on a normal distribution or a lognormal distribution. Many studies on location information are based on the fact that GPS location measurement errors follow a normal distribution [30-33]. This study [34] showed the distributions that describe navigation positioning system errors more accurately include lognormal distributions. Therefore, the experiments were conducted by assuming that the measurement errors were based on a normal distribution and a lognormal distribution.

In the case wherein the measurement error is the lognormal distribution, experiments are only compared to *Geo-I*. As TDP is based on the case where the measurement error is the normal distribution, evaluations using TDP cannot be performed for the lognormal distribution.

As mentioned in the proposed method, the final noise is a combination of measurement errors and noise due to the Laplace mechanism. In the simulation, as shown in Fig. 1, the noise vector of measurement error due to the normal distribution and the noise vector due to the Laplace mechanism to satisfy differential privacy were added and used as the total noise.

The noise average and mean square error (MSE) are summarized in the results. Errors include both noise from the Laplace distribution for differential privacy and noise from the normal distribution as measurement errors.

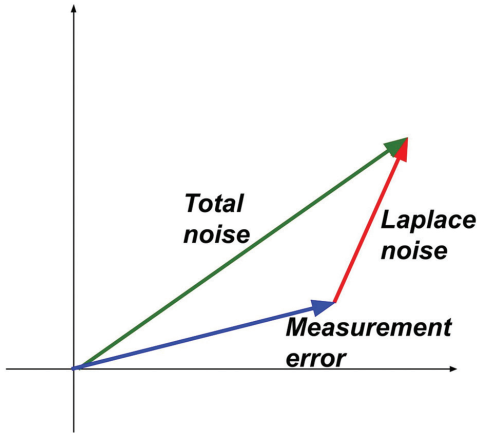

Fig. 1. Composition of angles

4.2. SIMULATION PARAMETERS

To find threshold w, the distance between x and x' Δ as $d(x, x')$, we conducted experiments with 1.0. The width of a histogram c is 0.5. Multiple of w to verify α is 0.5. The scope of verifying differential privacy δ is 10^{-3}. This means that we guarantee (ε, 10^{-3})-differential privacy. The number of samples ns is 10^8.

The measurement error was calculated from 2 types. One is a two-dimensional normal distribution with a standard deviation of 1.0. The other is the error by the radius from a lognormal distribution with a standard deviation of 1.0 and the angle is from a uniform distribution $[0, 2\pi)$. It is also used by MeasurementNoise() in Algorithm 2. For the noise generated from the Laplace distribution, we conducted experiments with $\epsilon = 1, 2, 5$, and 10.

In the numerical simulation, the number of samples is 10^8. In the Siafu simulation, the number of samples is 10^4. The space was divided into 500 × 500 squares, totaling 2,500 squares, and the noise average and noise MSE were calculated.

4.3. SIMULATION METHOD FOR TDP

TDP is focused on one-dimensional data basically [23]. In the method based on TDP, we consider x and y to be two independent variables.

According to Section 2.9., we generated noise with half the value ϵ and added it to x and y. For example, by adding noise generated from the Laplace distribution with $\varepsilon = 0.5$ for x and $\varepsilon = 0.5$ for y, we achieved total privacy protection of $\varepsilon = 1.0$.

5. EXPERIMENT RESULTS

In the case wherein the measurement error is the normal distribution, the total noise average and MSE of the numerical simulation are summarized in Tables 1 and 2, respectively. In the case wherein the measurement error is the normal distribution, the total noise average and MSE of the Siafu simulation are summarized in Tables 3 and 4, respectively. In the case wherein the measurement error is the normal distribution, the total noise average and MSE of the numerical simulation are summarized in Tables 5 and 6, respectively.

The total noise contains both Laplace noise for differential privacy and noise from the normal distribution as measurement errors. The results of the average amount of noise added to achieve differential privacy are summarized in Figs. 2, 3, and 4. When ε is close to 0, the noise is large.

According to all results, the proposed method has the smallest noise average and MSE compared with the other methods.

In the case wherein the measurement error is the normal distribution, the proposed method based on *T-Geo-I* reduced the noise average by up to 18% and 41% compared with methods based on *Geo-I* and TDP with numerical simulation, respectively. The proposed method based on *T-Geo-I* reduced the noise average by up to 15% and 36% compared with methods based on *Geo-I* and TDP with the Siafu simulation, respectively. The proposed method based on *T-Geo-I* reduced the noise MSE by up to 31% and 63% compared with *Geo-I* and TDP with numerical simulation, respectively. The proposed method based on *T-Geo-I* reduced the noise MSE by up to 17% and 38% compared with methods based on *Geo-I* and TDP with the Siafu simulation, respectively. The maximum reduction rate was achieved when $\varepsilon = 1, 2$.

In the case wherein the measurement error distribution is the lognormal distribution, the proposed *T-Geo-I* reduced the noise average and MSE by up to 60% and 67%, respectively, compared with *Geo-I* with numerical simulation. The maximum reduction rate was achieved when $\varepsilon = 1$.

In the case of $\epsilon = 5$ and 10, the result indicates that differential privacy is satisfied with only the measurement error without any noise addition because of the Laplace distribution. When $\epsilon = 10$, the noise averages of methods based on *T-Geo-I* and TDP are almost the same. This indicates that both methods do not add nearly any Laplace noise because differential privacy is almost satisfied with only the standard deviation when $\epsilon = 10$.

The proposed method can reduce the average amount of noise and is expected to enhance the usefulness of the data.

We tested them on a MacBook Air (M1, 2020), an Apple M1 CPU, and 16 GB of memory using Python. It takes

5 h to generate 10^8 Laplace noises. It takes 5 min to read the data of 10^8 Laplace noises already generated.

After the data are read, it takes 5 min for each value of w to create a histogram and verify whether differential privacy is satisfied.

We also experimented to see how much time it takes to protect privacy in a real environment. We measured the calculation time for acquiring location information and adding noise to the location information using an iPhone 13 mini. The average time value was calculated by measuring 100 times. The result is Fig. 5. The computation time for all methods was almost the same. Privacy protection can be achieved in a short time of 270-290 ms. This means that the proposed method is not algorithmically inefficient.

Table 1. Comparison of total noise average with numerical simulation (measurement error of normal distribution)

ϵ	*w* for *T-Geo-I*	*T-Geo-I* (noise average)	*Geo-I* (noise average)	TDP (noise average)
1	2.5	2.02	2.41	3.46
2	2.5	1.33	1.64	1.92
5	inf	1.25	1.33	1.27
10	inf	1.25	1.27	1.25

Table 2. Comparison of total noise MSE with numerical simulation (measurement error of normal distribution)

ϵ	*w* for *T-Geo-I*	*T-Geo-I* (noise MSE)	*Geo-I* (noise MSE)	TDP (noise MSE)
1	2.5	6.54	7.99	17.72
2	2.5	2.39	3.50	5.31
5	inf	1.99	2.23	2.07
10	inf	1.99	2.05	2.00

Table 3. Comparison of total noise average with Siafu simulation (measurement error of normal distribution).

ϵ	*w* for *T-Geo-I*	*T-Geo-I* (noise average)	*Geo-I* (noise average)	TDP (noise average)
1	2.5	1.96	2.28	3.09
2	2.5	1.40	1.65	1.84
5	inf	1.36	1.40	1.36
10	inf	1.36	1.37	1.36

Table 4. Comparison of total noise MSE with Siafu simulation (measurement error of normal distribution).

ϵ	*w* for *T-Geo-I*	*T-Geo-I* (noise MSE)	*Geo-I* (noise MSE)	TDP (noise MSE)
1	2.5	1.96	2.28	3.09
2	2.5	1.40	1.65	1.84
5	inf	1.36	1.40	1.36
10	inf	1.36	1.37	1.36

Table 5. Comparison of total noise average with numerical simulation (measurement error of lognormal distribution).

ϵ	*w* for *T-Geo-I*	*T-Geo-I* (noise average)	*Geo-I* (noise average)
1	inf	1.65	4.17
2	inf	1.65	3.74
5	inf	1.65	3.60
10	inf	1.65	3.58

Table 6. Comparison of total noise MSE with numerical simulation (measurement error of lognormal distribution).

ϵ	*w* for *T-Geo-I*	*T-Geo-I* (noise MSE)	*Geo-I* (noise MSE)
1	inf	7.39	22.85
2	inf	7.39	18.35
5	inf	7.39	17.09
10	inf	7.39	16.91

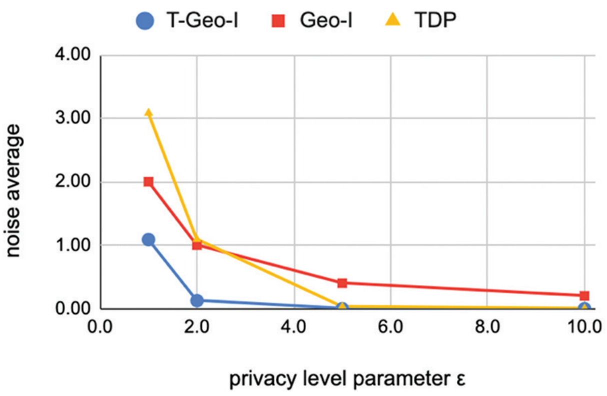

Fig. 2. Average amount of noise added to achieve differential privacy with numerical simulation (measurement error of normal distribution)

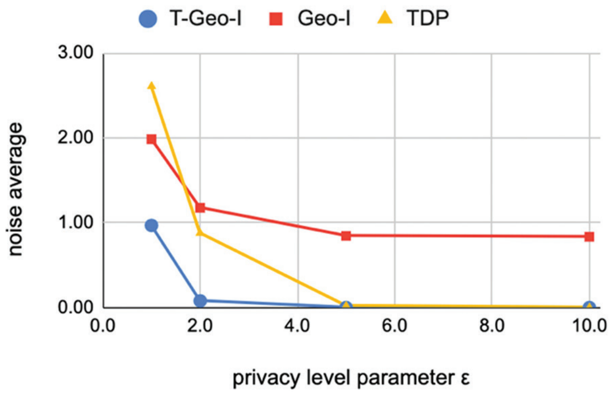

Fig. 3. Average amount of noise added to achieve differential privacy with Siafu simulation (measurement error of normal distribution)

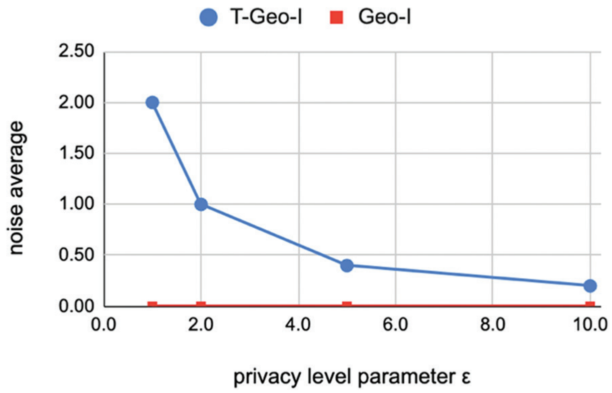

Fig. 4. Average amount of noise added to achieve differential privacy with numerical simulation (measurement error of lognormal distribution)

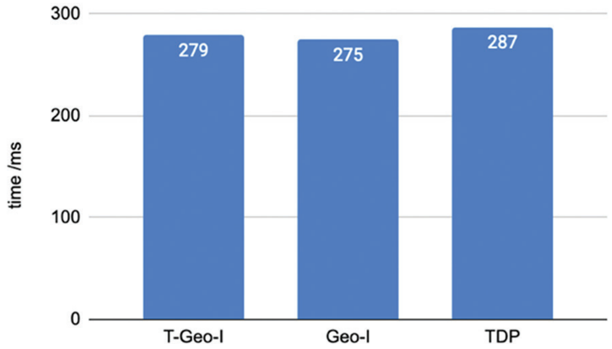

Fig. 5. The time required to measure the location information on the user's smartphone device and to apply differential privacy noise

6. DISCUSSION

In Apple's development, the privacy level parameter ϵ is equal to 1 or 2 per datum [35]. For example, Apple's differential privacy team used $\epsilon = 2$, 4, and 8 for their experiment evaluations [36]. In the study that proposed RAPPOR by Google, $\epsilon = \log(3)$ was used as the main parameter [37]. In TDP study [23], ϵ is set in the range 1–10. Therefore, we experimented with $\epsilon = 1$, 2, 5, and 10.

It was confirmed that the noise average was reduced not only in the numerical simulation but also in the Siafu simulation. The Siafu simulation is based on a typical human behavior model on a map. This means that we can expect to enhance the usefulness of data even in real-life situations.

As shown in Figs. 2, 3, and 4, the noise average was higher at a higher degree of privacy protection. This means that the effect of noise reduction using the proposed method is high if the degree of privacy protection is high. According to the results in Tables 1–6, the greatest reduction effect is obtained when ε = 1 and 2. Because the Laplace noise is small when ε = 5 and 10, the reduction in the total noise of the measurement error and the Laplace noise is small.

The case where $w=inf$ means that no Laplace noise is added. When $w=inf$, the noise regarding the proposed *T-Geo-I* is from only measurement error. In other words, $(\varepsilon,10^{-3})$-differential privacy is satisfied even without adding any Laplace noise. It is shown that there are cases wherein privacy can be protected using only measurement errors. Note that the proposed method satisfies differential privacy at the specified level. In other words, the existing methods add unnecessary noise beyond the specified level.

It takes more than 5 h to calculate the threshold w when the number of samples is 10^8. However, once the value of w is calculated, the determined value w can be repeatedly used for actual privacy protection. 10^8 Laplace noise generation is necessary for the simulation to determine the threshold w and only needs to be done once on the server side.

On the contrary, actual privacy protection takes a very short time. As shown in the newly added Fig. 5, actual privacy protection has a low computational cost. The computation time for all methods was almost the same. Privacy protection can be achieved in a short time of 270-290 ms. This shows that the proposed method is not algorithmically inefficient. Because this method has a very low computational cost, it can be easily introduced into various systems. The usefulness of the data can be improved compared with conventional methods.

The conventional method TDP assumes a normal distribution of measurement errors [23]. Our method is not limited to normal distributions. An appropriate threshold value w can be determined by simulation of any distribution. This is an advantage of our method.

In this experiment, we assumed a normal distribution and a lognormal distribution for measurement errors. Many studies have been conducted on measurement errors in location information. They are affected by various factors such as radio waves and weather conditions. They cannot be determined in one way. There is also research on simulation measurement errors [34, 38]. In the future, experiments are expected to be conducted on measurement errors in various situations.

The disadvantage is that the simulation for finding the threshold value w is computationally expensive. In the future, methods for determining the threshold value w based on the proof of mathematical formulas instead of simulation are expected.

Our method does not consider continuous location information. By acquiring continuous location information based on the trajectory of a person's movement, the risk of estimating the person's true location is increased [39–40]. In the future, we intend to address these issues.

7. CONCLUSION

Systems that collect location information and publish statistics, such as those that publish congestion information, have been extensively employed. These systems use differential privacy to ensure the privacy of user data. Privacy protection using the Laplace mechanism based on differential privacy adds noise, which reduces the usefulness of the data when the degree of privacy protection is high. Therefore, we focus on the fact that the values obtained by measurement devices contain errors and propose a location information privacy protection method that reduces the amount of added noise.

In the case wherein the measurement error is the normal distribution, the proposed method based on *T-Geo-I* succeeded in reducing the noise average by up to 18% and 41% compared with methods based on *Geo-I* and TDP, respectively, while maintaining a prespecified level of privacy in 10^8 samples of numerical data. It also reduced the noise MSE by up to 31% and 63% compared with methods based on *Geo-I* and TDP, respectively. The proposed method based on *T-Geo-I* reduced the noise average by up to 15% and 36% compared with methods based on *Geo-I* and TDP, respectively, in a location simulation of the human behavior of 10^4 users on a map using a typical human behavior model. It also reduced the noise MSE by up to 17% and 38% compared with methods based on *Geo-I* and TDP, respectively.

In the case wherein the measurement error is the lognormal distribution, the proposed method based on *T-Geo-I* succeeded in reducing the noise average and MSE by up to 60% and 67%, respectively, compared with methods based on *Geo-I*, while maintaining a prespecified level of privacy in 10^8 samples of numerical data.

The maximum reduction rate was achieved when ε is small: the privacy protection level high.

These findings demonstrate that our method can improve the usefulness of data while maintaining a prespecified privacy protection level.

8. REFERENCE

[1] T. Alam, B. Rababah, A. Ali, S. Qamar, "Distributed Intelligence at the Edge on IoT Networks", Annals of Emerging Technologies in Computing, Vol. 4, No. 5, 2020, pp. 1-18.

[2] G. Muneeswari, A. Ahilan, R. Rajeshwari, K. Kannan, C. J. C. Singh, "Trust and Energy-Aware Routing Protocol for Wireless Sensor Networks Based on Secure Routing", International Journal of Electrical and Computer Engineering Systems, Vol. 14, No. 9, 2023, pp. 1015-1022.

[3] D. Liu, X. Gao, H. Wang, "Location privacy breach: Apps are watching you in background", Proceedings of the IEEE 37th International Conference on Distributed Computing Systems, Atlanta, GA, USA, 5-8 June 2017, pp. 2423-2429.

[4] S. Kumar et al. "Protecting location privacy in cloud services", Journal of Discrete Mathematical Sciences and Cryptography, Vol. 25, No. 4, 2022, pp. 1053-1062.

[5] K. S. Saraswathy, S. S. Sujatha, "Using Attribute-Based Access Control, Efficient Data Access in the Cloud with Authorized Search" International Journal of Electrical and Computer Engineering Systems, Vol. 13, No. 7, 2022, pp. 569-575.

[6] G. Sun et al. "Location Privacy Preservation for Mobile Users in Location-Based Services", IEEE Access, Vol. 7, 2019, pp. 87425-87438.

[7] N. Ahmed, Z. Deng, I. Memon, F. Hassan, K. H. Mohammadani, R. Iqbal, "A Survey on Location Privacy Attacks and Prevention Deployed with IoT in Vehicular Networks", Wireless Communications and Mobile Computing, Vol. 2022, 2022.

[8] E. P. de Mattos, A. C. S. A. Domingues, B. P. Santos, H. S. Ramos, A. A. F. Loureiro, "The Impact of Mobility on Location Privacy: A Perspective on Smart Mobility", IEEE Systems Journal, Vol. 16, No. 4, 2022 pp. 5509-5520.

[9] S. Özdal Oktay, S. Heitmann, C. Kray, "Linking location privacy, digital sovereignty and location-based services: a meta review", Journal of Location Based Services, Vol. 18, No. 1, 2024, pp. 1-52.

[10] M. K. Gupta, A. K. Rai, B. Pandey, A. Gupta, V. K. Verma, "Big Data Privacy: A Survey Paper", Proceedings of the International Conference on IoT, Communication and Automation Technology, Gorakhpur, India, 2023, pp. 1-6.

[11] A. Fathalizadeh, V. Moghtadaiee, M. Alishahi, "Indoor Geo-Indistinguishability: Adopting Differential Privacy for Indoor Location Data Protection", IEEE Transactions on Emerging Topics in Computing, 2023. (in press)

[12] P. Zhang, X. Cheng, S. Su, N. Wang, "Task Allocation Under Geo-Indistinguishability via Group-Based Noise Addition", IEEE Transactions on Big Data, Vol. 9, No. 3, 2023, pp. 860-877.

[13] E. T. Martínez Beltrán et al. "Decentralized Federated Learning: Fundamentals, State of the Art, Frameworks, Trends, and Challenges", Proceedings of the IEEE Communications Surveys & Tutorials, Vol. 25, No. 4, 2023, pp. 2983-3013.

[14] C. Dwork, “Differential Privacy”, Proceedings of the 33rd International Colloquium on Automata, Languages, and Programming, Vencie, Italy, 10-14 July 2006, pp. 1-12.

[15] C. Dwork, “Differential Privacy: A Survey of Results”, Proceedings of the International Conference on Theory and Applications of Models of Computation, Xi'an, China, 25-29 April 2008, pp. 1-19.

[16] C. Dwork, A. Roth, “The algorithmic foundations of differential privacy”, Foundations and Trends in Theoretical Computer Science, Vol. 9, No. 3-4, 2014, pp. 211-407.

[17] C. Dwork, F. McSherry, K. Nissim, A. Smith, “Calibrating noise to sensitivity in private data analysis”, Proceedings of the Theory of Cryptography: Third Theory of Cryptography Conference, New York, NY, USA, 4-7 March 2006, pp. 265-284.

[18] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, M. Naor, “Our data, ourselves: Privacy via distributed noise generation”, Advances in Cryptology – EUROCRYPT 2006, Proceedings of the 25th Annual International Conference on the Theory and Applications of Cryptographic Techniques, St. Petersburg, Russia, 28 May - 1 June 2006, pp. 486-503.

[19] R. Dewri, “Local differential perturbations: Location privacy under approximate knowledge attackers”, IEEE Transactions on Mobile Computing, Vol. 12, No. 12, 2012, pp. 2360-2372.

[20] J. C. Duchi, M. I. Jordan, M. J. Wainwright, “Local privacy and statistical minimax rates”, Proceedings of the IEEE 54th Annual Symposium on Foundations of Computer Science, Berkeley, CA, USA, 26-29 October 2013, pp. 429-438.

[21] K. Chatzikokolakis, M. E. Andrés, N. E. Bordenabe, C. Palamidessi, “Broadening the Scope of Differential Privacy Using Metrics”, Proceedings of Privacy Enhancing Technologies: 13th International Symposium, Bloomington, IN, USA, 10-12 July 2013, pp. 82-102.

[22] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, C. Palamidessi, “Geo-indistinguishability: Differential privacy for location-based systems”, Proceedings of the ACM SIGSAC Conference on Computer & Communications Security, November 2013, pp. 901-914.

[23] Y. Sei, A. Ohsuga, “Private true data mining: Differential privacy featuring errors to manage Internet-of-Things data”, IEEE Access, Vol. 10, 2022, pp. 8738-8757.

[24] P. Kairouz, S. Oh, P. Viswanath, “The composition theorem for differential privacy”, Proceedings of the 23rd International Conference on Machine Learning, Lille, France, 2015, pp. 1376-1385.

[25] J. Konečný, H. B. McMahan, D. Ramage, P. Richtárik, “Federated optimization: Distributed machine learning for on-device intelligence”, arXiv:1610.02527, 2016.

[26] Y. Zhang, Y. Lu, F. Liu, “A systematic survey for differential privacy techniques in federated learning”, Journal of Information Security, Vol. 14, No. 2, 2023, pp. 111-135.

[27] Z. Zong, M. Yang, J. Ley, A. Markopoulou, C. Butts, “Privacy by Projection: Federated Population Density Estimation by Projecting on Random Features”, Proceedings on Privacy Enhancing Technologies, Vol. 2023, No. 1, 2023, pp. 309-324.

[28] M. Martin, P. Nurmi, “A Generic Large Scale Simulator for Ubiquitous Computing”, Proceedings of the Third Annual International Conference on Mobile and Ubiquitous Systems: Networking & Services, San Jose, CA, USA 17-21 July 2006, pp. 1-3.

[29] Y. Sei, A. Ohsuga, “Location Anonymization With Considering Errors and Existence Probability”, IEEE Transactions on Systems, Man, and Cybernetics: Systems, Vol. 47, No. 12, 2016, pp. 3207-3218.

[30] P. Chao, W. Hua, R. Mao, J. Xu, X. Zhou, “A Survey and Quantitative Study on Map Inference Algorithms From GPS Trajectories”, IEEE Transactions on Knowledge and Data Engineering, Vol. 34, No. 1, 2020, pp. 15-28.

[31] E. Frentzos, K. Gratsias, Y. Theodoridis, “On the Effect of Location Uncertainty in Spatial Querying”, IEEE transactions on Knowledge and Data Engineering, Vol. 21, No. 3, 2008, pp. 366-383.

[32] D. Zhang, Z. Chang, S. Wu, Y. Yuan, K.-L. Tan, G. Chen, “Continuous Trajectory Similarity Search for Online Outlier Detection”, IEEE Transactions on Knowledge and Data Engineering, Vol. 34, No. 10, 2020, pp. 4690-4704.

[33] T. Ogino, “GPS Improvement System Using Short-Range Communication”, Proceedings of the International Conference on Computing, Networking and Communications, Maui, HI, USA, 5-8 March 2018, pp. 82-87.

[34] M. Specht, “Consistency of the Empirical Distributions of Navigation Positioning System Errors with Theoretical Distributions—Comparative Analysis of the DGPS and EGNOS Systems in the Years 2006 and 2014”, Sensors, Vol. 21, No. 1, 2020, p. 31.

[35] J. Tang, A. Korolova, X. Bai, X. Wang, X. Wang, “Privacy Loss in Apple's Implementation of Differential Privacy on MacOS 10.12”, arXiv:1709.02753, 2017

[36] Differential Privacy Team, “Learning with privacy at scale”, Apple Machine Learning Research, Apple, December 2017.

[37] Ú. Erlingsson, V. Pihur, A. Korolova, “RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response”, Proceedings of the ACM SIGSAC Conference on Computer and Communications Security, November 2014, pp. 1054-1067.

[38] A. El Abbous, N. Samanta, “A modeling of GPS error distributions”, Proceedings of the European Navigation Conference, Lausanne, Switzerland, 9-12 May 2017. pp. 119-127.

[39] Y. Zhao, J. Chen, “Vector-Indistinguishability: Location Dependency Based Privacy Protection for Successive Location Data”, IEEE Transactions on Computers, 2023. (in press)

[40] X. Sun et al. “Synthesizing Realistic Trajectory Data with Differential Privacy”, IEEE Transactions on Intelligent Transportation Systems, Vol. 24, No. 5, 2023, pp. 5502-5515.